

ORTHOGONAL EXPANSION OF QUANTILE FUNCTION
AND COMPONENTS OF THE SHAPIRO-FRANCIA STATISTIC

TECHNICAL REPORT NO. 8

AKIMICHI TAKEMURA

APRIL 1983

U. S. ARMY RESEARCH OFFICE
CONTRACT DAAG29-82-K-0156

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

ORTHOGONAL EXPANSION OF QUANTILE FUNCTION
AND COMPONENTS OF THE SHAPIRO-FRANCIA STATISTIC

TECHNICAL REPORT NO. 8

AKIMICHI TAKEMURA
STANFORD UNIVERSITY

APRIL 1983

U. S. ARMY RESEARCH OFFICE

CONTRACT DAAG 29-82-K-0156

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

§1 Introduction.

Let $F(x)$ be a distribution function on the real line with finite second order moment. Then the quantile function $F^{-1}(u)$ is square integrable as a function of $u \in (0, 1)$. Hence $F^{-1}(u)$ can be expanded into a Fourier series with respect to a complete orthonormal system of $L^2[0, 1]$: the space of square integrable functions on $(0, 1)$. This idea was used by Sugiura (1962, 1964) in connection with approximation of expected values of order statistics. In this paper we first note that various complete orthonormal systems of $L^2[0, 1]$ can be obtained by quantile transformation of complete orthonormal systems of $L^2(F)$ where $L^2(F)$ denotes the space of square integrable functions with respect to a distribution F , for example the normal distribution. If we take the normal distribution as F , then the Fourier coefficients can be thought as components with respect to an infinite dimensional coordinate system whose origin is the normal distribution itself. This gives a very crisp idea of distances and directions between various distributions. Furthermore the Fourier series expansion of quantile function closely parallels the Cornish-Fisher expansion (Fisher and Cornish, 1960) and gives a general method of approximating quantiles of one distribution based on quantiles of another distribution. The Cornish-Fisher expansion is asymptotic in its nature, whereas our expansion leads to a convergent series. These ideas are discussed in Section 2.

When F is taken to be an empirical distribution function F_n , the Fourier coefficients of F_n^{-1} provide a decomposition of the sum of squares constituting the Shapiro-Francia statistic (1972). This is analogous to the components of the Cramér-von Mises statistic discussed by Durbin and Knott (1972) and Durbin, Knott, and Taylor (1975). See also Durbin (1973). As in the case of components of the Cramér-von Mises statistic, various tests can be constructed from the components depending on the possible alternative hypotheses, whereas the Shapiro-Francia test serves as a test which detects an overall departure from the null hypothesis. The Shapiro-Francia statistic and its components can be applied to nonnormal distributions as well. See Section 3 for these developments.

In the study of asymptotic distributions of the Shapiro-Francia test statistic, de Wet and Venter (1972) implicitly used our components as eigenfunctions of the asymptotic covariance function of the quantile process. In the normal case, the eigenfunctions are the

Hermite polynomials. This has been noticed in other contexts too. See Stephens (1975). Despite this close analogy to the components of the Cramér-von Mises statistic, components of the Shapiro-Francia statistic have not been taken up as useful tools in themselves. We have to add, though, that components which are closely related to ours have been discussed in the framework of weighted Cramér-von Mises tests with various weight functions. See de Wet and Venter (1973a,b), Pettitt (1977,1978). This connection is based on the asymptotic equivalence between the Shapiro-Francia statistic and certain weighted Cramér-von Mises statistic. See Gregory (1977a,b). We give a brief discussion on this point in Section 4.

§2 Orthogonal expansion of quantile function.

2.1 Mallow's d_2 -distance and an angle between two distributions.

Let Γ_2 be the set of (right continuous) distribution functions on the real line having finite second order moments, $\Gamma_2 = \{F : \int x^2 dF < \infty\}$. Let F^{-1} be the quantile function which we take to be right continuous for definiteness: $F^{-1}(u) = \sup\{x : F(x) \leq u\}$. For $F \in \Gamma_2$ we have $\int_{-\infty}^{\infty} x^2 dF = \int_0^1 (F^{-1}(u))^2 du < \infty$. Hence by taking the quantile function Γ_2 is mapped into $L^2[0,1]$, the space of square integrable functions on $(0,1)$. Mallow's d_2 -distance between F_1 and F_2 in Γ_2 is defined by the usual inner product \langle, \rangle of $L^2[0,1]$, namely

$$(2.1) \quad d_2(F_1, F_2) = \left\{ \int_0^1 (F_1^{-1}(u) - F_2^{-1}(u))^2 du \right\}^{1/2}.$$

A nice discussion on this is given in Section 8 of Bickel and Freedman (1981).

Proposition 2.1. *Let $F_1, F_2 \in \Gamma_2$. Then $d_2(F_1, F_2) = 0$ if and only if $F_1 = F_2$.*

Proof is easy and omitted. From now on we are mainly interested in the cosine of angle between two distributions obtained from d_2 -distance. Let U be uniformly distributed on $(0,1)$. We define

$$(2.2) \quad \rho(F_1, F_2) = \text{Cor}(F_1^{-1}(U), F_2^{-1}(U)).$$

Note that ρ is invariant with respect to location and scale changes in F_1 and F_2 . For example if $F_1(x) = F_0((x-a)/b) = u$ ($b > 0$) then $F_1^{-1}(u) = a + bF_0^{-1}(u)$. Hence $\rho(F_1, F_2) = \text{Cor}(a + bF_0^{-1}(U), F_2^{-1}(U)) = \text{Cor}(F_0^{-1}(U), F_2^{-1}(U)) = \rho(F_0, F_2)$. Similarly for F_2 . Therefore ρ can be interpreted as the cosine of angle between location scale families. Also note that ρ is nonnegative because F_1^{-1}, F_2^{-1} are both monotone increasing.

Proposition 2.2. $\rho(F_1, F_2) = 1$ if and only if F_1 is a location scale transform of F_2 .

This follows from the fact that if F_1, F_2 have mean 0 and variance 1, then $d_2(F_1, F_2)^2 = 2(1 - \rho(F_1, F_2))$.

2.2 Orthonormal basis with respect to a specific distribution.

By taking the quantile functions we mapped Γ_2 into $L^2[0, 1]$. Now we take an appropriate complete orthonormal system in $L^2[0, 1]$. The important point is that this can be done with respect to a specific distribution. First we prove

Proposition 2.3. Let $F \in \Gamma_2$ be a continuous distribution function which is supported (i.e., strictly increasing) on an interval. Then $\{\psi_i\}_{i=0}^\infty$ is a complete orthonormal system of $L^2(F) = \{\phi : \int \phi(x)^2 dF(x) < \infty\}$ if and only if $\{\psi_i \circ F^{-1}\}_{i=0}^\infty$ is a complete orthonormal system of $L^2[0, 1]$.

Proof: Let $\{\psi_i\}_{i=0}^\infty$ be a complete orthonormal system of $L^2(F)$. Then

$$\int_0^1 \psi_i(F^{-1}(u))\psi_j(F^{-1}(u))du = \int_{-\infty}^\infty \psi_i(x)\psi_j(x)dF(x) = \delta_{ij}.$$

Hence $\psi_i \circ F^{-1}$'s are orthonormal. To prove completeness suppose $f \in L^2[0, 1]$ and $\langle f, \psi_i \circ F^{-1} \rangle = 0$, $i = 0, 1, \dots$. By assumption $F(F^{-1}(u)) = u$. This implies that $f \circ F \in L^2(F)$ and $0 = \langle f, \psi_i \circ F^{-1} \rangle = \int f(F(x))\psi_i(x)dF$, $i = 0, 1, \dots$. Hence by completeness of $\{\psi_i\}_{i=0}^\infty$, $f \circ F = 0$ in $L^2(F)$ or $\int_{-\infty}^\infty f(F(x))^2 dF(x) = 0$. Hence $\int_0^1 f(u)^2 du = 0$. This shows that $\{\psi_i \circ F^{-1}\}_{i=0}^\infty$ is complete in $L^2[0, 1]$. Converse can be proved similarly. ■

We can now express G^{-1} in a Fourier series using the orthonormal basis $\{\psi_i \circ F^{-1}\}_{i=0}^\infty$. Since we are working with a location scale family, we take F with mean

0 and variance 1, namely $\int x dF = 0$, $\int x^2 dF = 1$. In this case by Gram-Schmidt orthonormalization we can take an orthonormal system $\{\psi_i\}_{i=0}^{\infty}$ of $L^2(F)$ with $\psi_0 \equiv 1$, $\psi_1 = x$, since these two are orthonormal. In the sequel we work with this choice of orthonormal basis.

Theorem 2.1. *Let $F \in \Gamma_2$ be a continuous distribution function supported on an interval and with mean 0 and variance 1. Let $\{\psi_i\}_{i=0}^{\infty}$ be a complete orthonormal system of $L^2(F) = \{\phi : \int \phi^2 dF < \infty\}$ with $\psi_0 \equiv 1$, $\psi_1 = x$. Then for any $G \in \Gamma_2$*

$$(2.3) \quad G^{-1}(u) = \sum_{i=0}^{\infty} a_i \psi_i(F^{-1}(u)),$$

where $a_i = \langle G^{-1}, \psi_i \circ F^{-1} \rangle$ and convergence on the right hand side is in $L^2[0, 1]$. Furthermore $a_i = 0$, $i = 2, 3, \dots$, if and only if G is a location scale transform of F .

Proof: We need only to prove the last statement. Let $\mu_G = \int x dG$. Note

$$\begin{aligned} a_0 &= \langle G^{-1}, 1 \rangle = \int_0^1 G^{-1}(u) du = \mu_G, \\ a_1 &= \langle G^{-1}, F^{-1} \rangle = \langle G^{-1} - \mu_G, G^{-1} - \mu_G \rangle^{1/2} \rho(G, F), \\ \langle G^{-1} - \mu_G, G^{-1} - \mu_G \rangle &= \sum_{i=1}^{\infty} a_i^2. \end{aligned}$$

Hence

$$(2.4) \quad \rho(G, F) = a_1 / \left(\sum_{i=1}^{\infty} a_i^2 \right)^{1/2}.$$

Therefore $\rho(G, F) = 1$ if and only if $a_i = 0$, $i = 2, 3, \dots$. ■

If we normalize G as well to have mean 0 and variance 1, we have $a_0 = \mu_G = 0$ and $\sum_{i=1}^{\infty} a_i^2 = 1$. Hence we can write

$$(2.5) \quad \rho(G, F) = a_1 = \left(1 - \sum_{i=2}^{\infty} a_i^2 \right)^{1/2}.$$

This shows that if we represent a location scale family of distributions by the normalized distribution G with mean 0 and variance 1, then it can be uniquely represented by

components a_i , $i = 2, 3, \dots$ with respect to (infinite dimensional) coordinate system $\{\psi_i \circ F^{-1}\}_{i=2}^{\infty}$. We call a_i 's *components of G with respect to F* . By Theorem 2.1 the origin in this coordinate system is F itself.

The components a_i , $i = 2, 3, \dots$, can be used in two ways. First it gives information on how much and in what direction G deviates from F . See Section 2.3 below. Second use is for an approximation of quantiles. Let y_α, x_α be α -quantiles of G and F , respectively. Then (2.3) implies

$$(2.6) \quad y_\alpha = \sum_{i=0}^{\infty} a_i \psi_i(x_\alpha).$$

Hence we can approximate α -quantile of G by (a function of) α -quantile of F . This is analogous to Cornish-Fisher expansion. See Fisher and Cornish (1960). Cornish-Fisher expansion is asymptotic in its nature, whereas (2.6) is applicable for any distribution with finite variance. Concerning the convergence of (2.6) the following distributional result is immediate from Theorem 2.1. Since L^2 -convergence implies convergence in distribution we have

Theorem 2.2. *Let X, Y be distributed according to F and G respectively. Let $X_n = \sum_{i=0}^n a_i \psi_i(X)$. Then $X_n \rightarrow Y$ in distribution as $n \rightarrow \infty$.*

Furthermore for standard distributions and standard orthonormal systems (2.6) converges pointwise. This is an obvious advantage of (2.6) over Cornish-Fisher expansion. More precisely we have

Theorem 2.3. *Let $G^{-1}(u)$ be piecewise continuously differentiable. If $F, \{\psi_i\}_{i=0}^{\infty}$ are either (i) normal distribution and the Hermite polynomials, or (ii) Gamma distribution and the Laguerre polynomials, or (iii) Beta distribution and the Jacobi polynomials, properly normalized, then the right hand side of (2.6) converges pointwise at every continuity point of G^{-1} . At a discontinuity point it converges to $(G^{-1}(u) + G^{-1}(u-0))/2$.*

These are standard results in the theory of orthogonal polynomials. See Chapter 4 of Lebedev (1965) or Chapter 9 of Szegő (1959). Various refinements of Theorem 2.3 can be found in these and other books on special functions and orthogonal polynomials.

2.3 Some examples.

We present components of several distributions with respect to normal, exponential, and uniform distributions. Here we consider distributions for which $G^{-1}(u)$ (or $F^{-1}(u)$) can be explicitly written and hence the components can be computed rather easily. More extensive tables of components are found in Takemura (1983).

	<i>unif</i>	<i>exp</i>	<i>logistic</i>	<i>weibull(2/3)</i>	<i>weibull(2)</i>	<i>beta(2, 2)</i>
	<i>wrt normal</i>	<i>wrt normal</i>	<i>wrt normal</i>	<i>wrt exp</i>	<i>wrt exp</i>	<i>wrt unif</i>
1	.9772	.9032	.9959	.9692	.9565	.9959
2		.4212(3.71)		.2423(3.22)	-.2391(32.8)	
3	-.1995(11.7)	.0818(.09)	.0902(.25)	-.0404(.53)	.1196(16.0)	.0851(11.3)
4		-.0116(.01)		.0151(.15)	-.0747(9.39)	
5	.0669(1.79)	-.0041(.00)	-.0045(.00)	-.0076(.06)	.0523(6.17)	.0264(2.72)
6				.0044(.03)	-.0392(4.36)	
7	-.0258(.31)			-.0028(.01)	.0308(3.25)	.0120(.95)
8				.0020(.00)	-.0251(2.51)	
9	.0106(.06)				.0209(2.00)	.0066(.41)
10					-.0177(1.63)	
11	-.0046(.01)				.0153(1.35)	.0041(.21)

There are six cases listed here. Each case has two columns: (i) a_k , (ii) $100 \times \sum_{i=k+1}^{\infty} a_i^2 / \sum_{i=2}^{\infty} a_i^2$, which is the remaining percentage out of $1 - \rho(G, F)^2$ when a_2, \dots, a_k are taken into account. The six cases are (i) uniform distribution with respect to normal, (ii) exponential distribution with respect to normal, (iii) logistic distribution with respect to normal, (iv) Weibull distribution with parameter .66 with respect to exponential, (v) Weibull distribution with parameter 2 with respect to exponential, (vi) Beta distribution with parameters (2,2) with respect to uniform, respectively. Orthonormal bases used are the Hermite, Laguerre, and Legendre polynomials for normal, exponential, and uniform distributions respectively. Details of computational procedures are found in Appendix.

§3 Components of the Shapiro-Francia statistic.

In the previous section distributions were thought as population distributions. In

this section we let $G = G_n$ be an empirical distribution function obtained from order statistics $X_{(1)} \leq \dots \leq X_{(n)}$. We show that the components of G_n give a decomposition of the Shapiro-Francia statistic: $W' = (\sum X_{(i)} m_i)^2 / (\sum (X_i - \bar{X})^2 \cdot \sum m_i^2)$ where $m_i = E(X_{(i)})$.

Note that $G_n^{-1}(u) = X_{(i)}$ for $(i-1)/n \leq u < i/n$. Hence we have

$$\begin{aligned}
 (3.1) \quad \langle G_n^{-1}, F^{-1} \rangle &= \sum_{i=1}^n X_{(i)} \int_{(i-1)/n}^{i/n} F^{-1}(u) du \\
 &= \sum_{i=1}^n X_{(i)} \int_{F^{-1}((i-1)/n)}^{F^{-1}(i/n)} x dF(x) \\
 &= \frac{1}{n} \sum_{i=1}^n X_{(i)} m'_i,
 \end{aligned}$$

where m'_i is the mean of the distribution which is obtained by truncating F from $(i-1)/n$ -quantile to i/n -quantile of F and renormalizing by n . Clearly m'_i is close to m_i . Therefore $\rho(G_n, F)^2$ is essentially the same as the Shapiro-Francia statistic (1972). See Shapiro and Wilk (1965) too. Actually $\rho(G_n, F)^2$ seems simpler than the Shapiro-Francia statistic because m'_i in (3.1) is generally easier to compute than the expected value of i -th order statistic. For the normal case see Section 3.1 below. Furthermore it is intuitively clear that $\rho(G_n, F) \rightarrow \rho(G, F)$ as $n \rightarrow \infty$ and by Proposition 2.2 the test based on $\rho(G_n, F)$ is consistent. This can be proved by simplifying and slightly modifying Sarkadi's proof (1975) of consistency of the Shapiro-Francia test.

Although $a_1 = \rho(G_n, F)$ gives a simple omnibus test we want to propose a larger class of tests which contains a_1 as a special case. As discussed in Section 2.2 we have the relation $a_1^2 = 1 - \sum_{i=2}^{\infty} a_i^2$. In view of this relation we consider tests based on the whole sequence (a_2, a_3, \dots) rather than based on a_1^2 alone. Since the components of G_n tend to be close to the components of the population distribution G , we can easily construct a test which is efficient against a particular alternative G . More specifically let G have components a_2^0, a_3^0, \dots . Now using the idea of regression let $T = \sum_{i=2}^{\infty} a_i a_i^0$. The test with the rejection region $T > c$ clearly has high power against the alternative G . As we can see from the examples of Section 2.3 components generally converge to zero fairly fast. Hence in $T = \sum_{i=2}^{\infty} a_i a_i^0$ and $\sum_{i=2}^{\infty} a_i^2$ the important terms are the first few terms. It is therefore essential to be able to interpret low order components.

Now we apply these ideas to tests of specific distributions.

3.1 Tests of normality.

Let $\phi(x) = \Phi'(x)$ be the standard normal density and $\{H_k(x)\}_{k=0}^{\infty}$ be the Hermite polynomials given in (A1) of Appendix. With respect to the orthonormal basis $\{H_k(\Phi^{-1}(u))/\sqrt{k!}\}$ components are given by $a_k = \langle G_n^{-1}, H_k \circ \Phi^{-1} \rangle / s\sqrt{k!}$ where $s^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / n$.

Now

$$\begin{aligned} s\sqrt{k!}a_k &= \sum_{i=1}^n X_{(i)} \int_{\Phi^{-1}((i-1)/n)}^{\Phi^{-1}(i/n)} H_k(x)\phi(x)dx \\ &= \sum_{i=1}^n X_{(i)} [H_{k-1}(Z_{(i-1)/n})\phi(Z_{(i-1)/n}) - H_{k-1}(Z_{i/n})\phi(Z_{i/n})], \end{aligned}$$

where $Z_\alpha = \Phi^{-1}(\alpha)$ is α -quantile of the standard normal distribution. We see that a_k 's are very simple to calculate. In particular $a_1 = \rho(G_n, \Phi) = \sum_{i=1}^n X_{(i)}(\phi(Z_{(i-1)/n}) - \phi(Z_{i/n}))/s$ corresponds to the Shapiro-Francia statistic.

In order to interpret low order terms let us look at them as regression coefficients in the following setting. Consider $G_n^{-1} \circ \Phi$. Plotting this function corresponds to plotting the observations on probability paper, where the normal scores are taken as x -axis. We give the weight $\phi(x)$ to the x -axis and consider the following weighted least square problem:

$$\min_{b_1, \dots, b_n} \int_{-\infty}^{\infty} (G_n^{-1}(\Phi(x)) - \sum_{k=0}^n b_k H_k(x)/\sqrt{k!})^2 \phi(x) dx.$$

Namely we approximate $G_n^{-1} \circ \Phi$ by a polynomial in the sense of weighted least squares. Because of the orthogonality of Hermite polynomials, the minimizing values of b_0, \dots, b_n are equal to the components a_0, \dots, a_n . Note that $H_2(x) = x^2 - 1$ and $H_3(x) = x^3 - 3x$. From the forms of these polynomials it is clear that a_2 roughly corresponds to skewness and a_3 roughly corresponds to kurtosis. Positive a_2 indicates skewness toward the right and positive a_3 indicates a heavier tail than the normal distribution. Another way of seeing this is the following. Consider $s\sqrt{2}a_2$. It can be approximated by $(1/n) \sum X_{(i)}(m_i^2 - 1)$ where $m_i = E(X_{(i)})$. This corresponds to $(1/n) \sum (X_{(i)}^3 - X_{(i)}) = \sum (X_i^3 - X_i)$. Similar consideration applies to higher order components. Examples in Section 2.3 confirm the above interpretations.

For *symmetric alternatives*, $T = \sum_{k=1}^{\infty} a_{2k+1}^2 / \sum_{k=0}^{\infty} a_{2k+1}^2 = 1 - a_1^2 / \sum_{k=0}^{\infty} a_{2k+1}^2$ is of considerable interest. Actually $T = 1 - \rho(G_s, \Phi)^2$ where G_s is such that $\rho(G_n, G_s) = \max\{\rho(G_n, G) \mid G \text{ is symmetric}\}$. The maximizing G_s can be given explicitly as

$$(3.2) \quad G_s^{-1}(u) = \frac{1}{2}(G_n^{-1}(u) - G_n^{-1}(1-u)).$$

By symmetry we have $\langle G_s^{-1}, H_{2k} \circ \Phi^{-1} \rangle = 0$, $\langle G_s^{-1}, H_{2k+1} \circ \Phi^{-1} \rangle = \langle G_n^{-1}, H_{2k+1} \circ \Phi^{-1} \rangle = a_{2k+1}$. Note that $s^2 \rho(G_n, G_s)^2 = \sum_{k=0}^{\infty} a_{2k+1}^2$. By Cauchy-Schwarz it is straightforward to show that this G_s maximizes $\rho(G_n, G)$ where G is symmetric. Furthermore $1 - \rho(G_s, \Phi)^2 = 1 - a_1^2 / \sum_{k=0}^{\infty} a_{2k+1}^2$ as claimed above. The denominator $\sum_{k=0}^{\infty} a_{2k+1}^2$ is the variance of G_s which can be calculated as follows. Note that G_s is the empirical distribution function of $(X_{(1)} - X_{(n)})/2, (X_{(2)} - X_{(n-1)})/2, \dots, (X_{(n)} - X_{(1)})/2$. Therefore $\sum_{k=0}^{\infty} a_{2k+1}^2 = \sum_{i=1}^n (X_{(i)} - X_{(n-i+1)})^2 / 4n$. Thus T can be computed easily.

We now turn to the asymptotic null distribution of each component. In the case of normality we have the following nice result:

Theorem 3.1. *Under the null hypothesis of normality $(\sqrt{n}a_{i_1}, \dots, \sqrt{n}a_{i_k})$, $2 \leq i_1 < \dots < i_k$ are asymptotically independently distributed according to the normal distribution with mean zero and variance $1/(i_j + 1)$, $j = 1, \dots, k$, respectively.*

This can be proved using results concerning the asymptotic normality of linear combinations of order statistics. For example the conditions of Theorem 1 of Shorack (1972) are straightforward to check. See Example 1b of Shorack (1972) in particular. Asymptotic independence of components follows from the formula for the asymptotic covariance in Corollary 4 of Chernoff, Gastwirth, and Johns (1967). Similarly under mild regularity conditions on G it is straightforward to show that $n^{1/2}T = n^{1/2} \sum_{i=2}^{\infty} a_i a_i^0$ has an asymptotic normal distribution with mean zero and variance

$$\begin{aligned} \sigma^2 &= \int_0^1 \int_0^1 (\min(s, t) - st) J(s) J(t) / [\phi(\Phi^{-1}(s)) \phi(\Phi^{-1}(t))] ds dt \\ &= \sum_{i=2}^{\infty} (a_i^0)^2 / (i + 1), \end{aligned}$$

where $\{a_i^0\}$ are components of G and $J(u) = G^{-1}(u) - a_1^0 \Phi^{-1}(u)$. See Theorem 3 of Chernoff, Gastwirth, and Johns (1967), Corollary 4.1 of Stigler (1969), and Theorem 1 of

Shorack (1972) for appropriate regularity conditions. As for the test against symmetric alternatives : $T = 1 - a_1^2 / \sum_{k=0}^{\infty} a_{2k+1}^2$, using results from de Wet and Venter (1972, 1973a) it can be shown that the distribution of $n(T - \mu_n)$ approaches the distribution of $(1/2) \sum_{k=1}^{\infty} (Y_k^2 - 1)/(k+1)$ where μ_n 's are suitable constants and Y_1, Y_2, \dots , are independent standard normal random variables. This is intuitively clear from Theorem 3.1.

3.2 Tests of exponentiality.

Let $f(x) = e^{-x}$, $F(x) = 1 - e^{-x}$ and $\{L_k(x)\}_{k=0}^{\infty}$ be the Laguerre polynomials given in (A1) of Appendix. In this case $F^{-1}(u) = -\log(1 - u)$ and

$$\int_{F^{-1}((i-1)/n)}^{F^{-1}(i/n)} L_k(x) f(x) = L_k^*[-\log(1 - (i-1)/n)] - L_k^*[-\log(1 - i/n)],$$

where $L_k^*(x) = (-1)^{n-1} (d^{n-1}/dx^{n-1})(x^n e^{-x})$. Similar considerations as in the normal case indicate that positive a_2 corresponds to a heavier tail toward the right than the exponential distribution. This is confirmed by examples in Section 2.3. Individual components and the test of the form $T = \sum_{i=2}^{\infty} a_i a_i^0$ can be used as in the normal case.

The asymptotic null distribution is not as nice as in the normal case. Again using results of Chernoff, Gastwirth, and Johns (1967) and others we have

Theorem 3.2. *Under the null hypothesis of exponentiality, $(\sqrt{n}a_{i_1}, \dots, \sqrt{n}a_{i_k})$, $2 \leq i_1 < \dots < i_k$ are asymptotically normally distributed with mean zero and covariance*

$$(3.3) \quad \sigma_{ij} = \int_0^{\infty} L_i^{**}(x) L_j^{**}(x) e^{-x} dx,$$

where $(i!)L_i^{**}(x) = \int_0^x L_i(y) dy - \int_0^{\infty} e^{-x} \int_0^x L_i(y) dy dx$.

$L_i^{**}(x)$ and σ_{ij} can be explicitly evaluated using (A1) in Appendix.

3.3 Tests of uniformity.

Here the results are similar to those for the exponential distribution. Let $\{P_k(x)\}_{k=0}^{\infty}$ be the Legendre polynomials given in (A1). Then $\{P_k(2u-1)\sqrt{2k+1}\}_{k=0}^{\infty}$ forms a complete orthonormal system on $[0,1]$. $\int_{(i-1)/n}^{i/n} P_k(2u-1) du$ can be explicitly evaluated using

(A1). Similar interpretations for low-order components can be given as in the normal and exponential cases. Under the null hypothesis $n^{1/2}a_i$'s are asymptotically normally distributed with mean zero and covariance

$$(3.4) \quad \sigma_{ij} = \int_0^1 P_i^*(x)P_j^*(x)dx,$$

where $P_i^*(x)/\sqrt{2i+1} = \int_0^x P_i(2u-1)du - \int_0^1 \int_0^x P_i(2u-1)dudx$. This can be explicitly evaluated using (A1).

§4 Discussion.

Here we discuss relations and relative advantages of our approach to other techniques especially the components of Cramér-von Mises statistics. Our approach is based on the empirical quantile function whereas the latter is based on the empirical distribution function.

One difficulty with techniques based on empirical distribution function is that it is not very natural to estimate parameters in that setting. When quantile function is used instead, estimation of location and scale parameters are taken care of by virtue of invariance, although actually this reduces to estimating mean and variance by their sample quantities. In fact the analysis becomes quite more complicated and individual cases have to be treated separately when various parameters are estimated. See for example Table 1 of Durbin, Knott, and Taylor (1975).

The reason that individual cases require separate treatments is that by "components" Durbin et al. mean components which are asymptotically independently distributed from each other. In our case this corresponds to obtaining principal components of the (infinite dimensional) asymptotic covariance matrices given in (3.3) and (3.4). Although this might be a logically natural thing to do on the ground that (i) components are uniquely determined, (ii) asymptotic distribution is understood better in this way, it might not be very useful from practical viewpoint. When components change depending on what parameter is estimated, it becomes harder to interpret them. The usefulness of the definition of our components is in Theorem 2.1 which is derived on the basis of population distributions rather than sampling properties and the interpretations of individual components seem to

be clearer. On the other hand our components do not possess the property of asymptotic independence in general. The case of normal null hypothesis is a fortunate exception.

A formal correspondence between $\int_0^1 (F_n^{-1}(u) - F^{-1}(u))^2 du$ and Cramér-von Mises type statistics is given as follows. $F_n^{-1}(u)$ can be written as $F^{-1}(\Gamma_n^{-1}(u))$ where Γ_n^{-1} is the quantile function obtained from observations from uniform distribution. Note that $F^{-1}(\Gamma_n^{-1}(u)) - F^{-1}(u) \approx (\Gamma_n^{-1}(u) - u)/f(F^{-1}(u))$ and $n^{1/2}(\Gamma_n^{-1}(u) - u)$ approaches a Brownian bridge as $n^{1/2}(\Gamma_n(u) - u)$ does. See Appendix in Shorack (1972) for these results. Therefore $\int_0^1 (F_n^{-1}(u) - F^{-1}(u))^2 du$ can be understood as Cramér-von Mises statistic where $1/f(F^{-1}(u))^2$ is taken as a weight function. See Gregory (1977a,b). Starting from the classical Anderson and Darling statistic (1952) various authors considered Cramér-von Mises statistics with various weight functions. See de Wet and Venter (1973a,b), Gregory (1977a,b), Pettitt (1977,1978). In particular, Pettitt (1977) considers the components of Cramér-von Mises statistic with the weight function $1/\phi(\Phi^{-1}(u))^2$ where $\phi = \Phi'$ is the standard normal density. From the above argument it is not surprising that there are close connections between his results and our results in Section 3.1.

Appendix: Computation of components.

We discuss computational procedures used to obtain components in Section 2.3.

Let

$$\begin{aligned}
 H_n(x) &= (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} = \sum_{k=0}^{[n/2]} \frac{(-1)^k n!}{2^k k! (n-2k)!} x^{n-2k} \\
 L_n(x) &= (-1)^n e^x \frac{d^n}{dx^n} e^{-x} x^n = \sum_{k=0}^n \frac{(-1)^k (n!)^2}{k! ((n-k)!)^2} x^{n-k} \\
 P_n(x) &= \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n = \sum_{k=0}^{[n/2]} \frac{(-1)^k (2n-2k)!}{2^n k! (n-k)! (n-2k)!} x^{n-2k}
 \end{aligned}
 \tag{A1}$$

be the Hermite, Laguerre, and Legendre polynomials, respectively. Then $\{H_n(x)/\sqrt{n!}\}_{n=0}^\infty$, $\{L_n(x)/n!\}_{n=0}^\infty$, $\{P_n(2x-1)\sqrt{2n+1}\}_{n=0}^\infty$, are complete orthonormal systems of L^2 -spaces associated with standard normal, standard exponential, and uniform distributions respectively.

For symmetric distributions we have the following result.

Proposition A1. *Let $G \in \Gamma_2$ be symmetric about its mean μ_G . Let F, ψ_k be either (i) standard normal and the Hermite polynomials, or (ii) uniform and the Legendre polynomials, suitably normalized. Then $a_{2k} = \langle G^{-1}, \psi_{2k} \circ F^{-1} \rangle = 0, k = 1, 2, \dots$*

This follows from the fact that $\psi_{2k} \circ F^{-1}$ is even about $u = 1/2$. For some cases the following lemma is useful.

Lemma A1. *Let F, G have positive densities f, g respectively. Let $h_n(x)f(x)$ be an indefinite integral of $\psi_n(x)f(x)$. Suppose that $\lim_{u \rightarrow 1,0} G^{-1}(u)f(F^{-1}(u))h_n(F^{-1}(u)) = 0$. Then*

$$a_n = - \int_0^1 \frac{f(F^{-1}(u))}{g(G^{-1}(u))} h_n(F^{-1}(u)) du.$$

Proof is by integration by parts.

Now we discuss each of 6 cases treated in Section 2.3.

Case 1. Uniform with respect to Normal. It is easy to obtain $a_1 = 1/(2\sqrt{\pi})$. Since $G(u) = u$ is bounded the assumption of Lemma A1 is clearly satisfied. Hence $\sqrt{n!}a_n = \int_0^1 H_{n-1}(\Phi^{-1}(u))\phi(\Phi^{-1}(u))du$. Using $H_k(x) = xH_{k-1}(x) - (k-1)H_{k-2}(x)$ we obtain the following recurrence relation $\sqrt{n!}a_n = -[(n-2)/2]\sqrt{(n-2)!}a_{n-2}$. This with a_1 above gives all components in view of Proposition A1.

Case 2. Exponential with respect to Normal. Again the condition of Lemma A1 is satisfied and

$$\sqrt{n!}a_n = \int_{-\infty}^{\infty} \frac{\phi(x)}{1 - \Phi(x)} H_{n-1}(x)\phi(x)dx.$$

This was evaluated by using Hermite-Gauss quadrature formula with 20 points.

Case 3. Logistic with respect to Normal. Given the components of exponential distribution, the components of logistic distribution can be obtained immediately. Let $G^{-1}(u) = -\log(1-u)$ be the quantile function of the exponential distribution. Then the quantile function of the Logistic distribution can be written as $(1/2)(G^{-1}(u) - G^{-1}(1-u))$ corresponding to G_s^{-1} in (3.2). Therefore as in Section 3.1 we see that the odd order components of logistic distribution are the same as those of exponential distribution up to a multiplicative constant. Even order components vanish by Proposition A1.

Cases 4 and 5. *Weibull with respect to Exponential.* In this case components can be explicitly written down. Let $G(x) = 1 - \exp(-x^c)$ be the distribution function of Weibull distribution with parameter c . It is the distribution of $Y^{1/c}$ where Y is a standard exponential random variable. Hence $G^{-1}(F(y)) = y^{1/c}$ and

$$\begin{aligned} n!a_n &= \int_0^\infty G^{-1}(F(y))L_n(y)e^{-y}dy \\ &= \int_0^\infty y^{1/c}L_n(y)e^{-y}dy \\ &= \sum_{i=0}^n c_i \Gamma(1/c + n - i + 1), \end{aligned}$$

where c_i 's are coefficients of the Laquerre polynomial given in (A1).

Case 6. *Beta(2,2) with respect to Uniform.* Here the beta distribution is taken to be on $[-1,1]$ for convenience. Let $g(x) = (3/4)(1 - x^2)$. Then $G(x) = (3/4)x - (1/4)x^3 + 1/2$. Therefore

$$\begin{aligned} \sqrt{n + \frac{1}{2}}a_n &= \int_0^1 G^{-1}(u)P_n(2u - 1)du \\ &= \int_{-1}^1 xP_n(2G(x) - 1)g(x)dx \\ &= \int_{-1}^1 xP_n\left(\frac{3}{2}x - \frac{1}{2}x^3\right)g(x)dx. \end{aligned}$$

This can be easily evaluated using (A1).

Acknowledgements. I wish to thank T.W. Anderson for his valuable suggestions. This research was supported in part by Army Research Office Contract DAAG29-82-K-0156.

References.

- Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann.Math.Statist.*, **23**, 193-212.
- Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann.Statist.*, **9**, 1196-1217.

- Chernoff, H., Gastwirth, J.L., and Johns, M.V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann.Math. Statist.*, **38**, 52-72.
- de Wet, T. and Venter, J.H. (1972). Asymptotic distributions of certain test criteria of normality. *S.Afr.Statist.J.*, **6**, 135-149.
- de Wet, T. and Venter, J.H. (1973a). Asymptotic distributions for quadratic forms with applications to tests of fit. *Ann.Statist.*, **1**, 380-387.
- de Wet, T. and Venter, J.H. (1973b). A goodness of fit test for a scale parameter family of distributions. *S.Afr.Statist.J.*, **7**, 35-46.
- Durbin, J. (1973). *Distribution theory for tests based on the sample distribution function*. SIAM, Philadelphia.
- Durbin, J. and Knott, M. (1972). Components of Cramér-von Mises statistics, I. *J.Roy.Statist. Soc. Ser.B*, **34**, 290-307.
- Durbin, J., Knott, M., and Taylor, C.C. (1975). Components of Cramér-von Mises statistics. II. *J.Roy.Statist.Soc. Ser.B*, **37**, 216-237.
- Fisher, R.A. and Cornish, E.A. (1960). The percentile points of distributions having known cumulants. *Technometrics*, **2**, 209-225.
- Gregory, G.G. (1977a). Large sample theory for U -statistics and tests of fit. *Ann.Statist.*, **5**, 110-123.
- Gregory, G.G. (1977b). Functions of order statistics and tests of fit. *S.Afr.Statist.J.*, **11**, 99-118.
- Lebedev, N.N. (1965). *Special functions and their applications*. Prentice-Hall, New Jersey.
- Pettitt, A.N. (1977). A Cramér-von Mises type goodness of fit statistic related to $\sqrt{b_1}$ and b_2 . *J.Roy.Statist.Soc. Ser.B*, **39**, 364-370.
- Pettitt, A.N. (1978). Generalized Cramér-von Mises statistics for the gamma distribution. *Biometrika*, **65**, 232-235.
- Sarkadi, K. (1975). The consistency of the Shapiro-Francia test. *Biometrika*, **62**, 445-450.

-
- Shapiro, S.S. and Francia, R.S. (1972). An approximate analysis of variance test for normality. *J.Amer.Statist.Assoc.*, **67**, 215-216.
- Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591-611.
- Shorack, G.R. (1972). Functions of order statistics. *Ann.Math.Statist*, **43**, 412-427.
- Stephens, M.A. (1975). Asymptotic properties for covariance matrices of order statistics. *Biometrika*, **62**, 23-28.
- Stigler, S.M. (1969). Linear functions of order statistics. *Ann.Math.Statist.*, **40**, 770-778.
- Sugiura, N. (1962). On the orthogonal inverse expansion with an application to the moments of order statistics. *Osaka Math.J.*, **14**, 253-63.
- Sugiura, N. (1964). The bivariate orthogonal inverse expansion and the moments of order statistics. *Osaka J.Math.*, **1**, 45-59.
- Szegő, G. (1959). *Orthogonal polynomials*. American Mathematical Society, Rhode Island.
- Takemura, A. (1983). Table of components of quantile functions. In preparation.

TECHNICAL REPORTS

U.S. ARMY RESEARCH OFFICE - CONTRACT DAAG29-82-K-0156

1. "Maximum Likelihood Estimators and Likelihood Ratio Criteria for Multivariate Elliptically Contoured Distributions," T.W. Anderson and Kai-Tai Fang, September 1982.
2. "A Review and Some Extensions of Takemura's Generalizations of Cochran's Theorem," George P.H. Styan, September 1982.
3. "Some Further Applications of Finite Difference Operators," Kai-Tai Fang, September 1982.
4. "Rank Additivity and Matrix Polynomials," George P.H. Styan and Akimichi Takemura, September 1982.
5. "The Problem of Selecting a Given Number of Representative Points in a Normal Population and a Generalized Mills' Ratio," Kai-Tai Fang and Shu-Dong He, October 1982.
6. "Tensor Analysis of ANOVA Decomposition," Akimichi Takemura, November 1982.
7. "A Statistical Approach to Zonal Polynomials," Akimichi Takemura, January 1983.
8. "Orthogonal Expansion of Quantile Function and Components of the Shapiro-Francia Statistics," Akimichi Takemura, April 1983.
9. "An Orthogonally Invariant Minimax Estimator of the Covariance Matrix of a Multivariate Normal Population," Akimichi Takemura, April 1983.

20. Abstract.

Let F be a distribution function on the real line having finite variance. Then the quantile function F^{-1} belongs to $L^2[0,1]$. Hence F^{-1} can be expanded in a Fourier series with respect to an appropriate complete orthonormal system of $L^2[0,1]$. When F is taken to be an empirical distribution function F_n , this leads to a decomposition of the Shapiro-Francia test statistic. This is analogous to the components of the Cramér-von Mises statistic discussed by Durbin et al. (1972,1975). The technique is applicable to a general distribution and tests of exponentiality and uniformity are discussed in addition to tests of normality.